

BUYING AI VOICE · 2026

Four Questions for Your AI Phone Vendor

The demonstrations are all excellent, because the conversational part is a commodity. These four questions cover everything the demonstration does not show.

Sit through four AI phone demonstrations in a week and they blur. The voices are equally natural, each handles an interruption, each gets the public holiday question right. The differences are entirely in what the system is **allowed to do**, what it does when it does not know, where the call goes when it cannot help, and which country your customers' conversations end up in.

■ Question 1. Can it write, or only read?

What you will hear	What it means	Verdict
"It sends your team an email."	Someone still enters it. The work moved to 8am tomorrow.	Reads only
"It creates a task in our portal."	Same, with a queue, plus a second place staff must look.	Reads only
"It logs the enquiry and notifies you."	Message taking with better handwriting.	Reads only
"Booking is in the calendar, customer has the SMS, note is on the contact."	Finished. Nobody touched it.	Writes

Question 2. What does it do when it does not know?

Three behaviours, one of them safe. **Says so plainly** and offers a callback, which is what a good new employee does in week one. **Deflects into a menu**, tolerable once, infuriating as a standard reply. **Answers anyway**, confidently and from general knowledge rather than your records. A system that will guess your opening hours will eventually guess your price. Ask how grounding is enforced. "The model is very accurate" is a hope, not an answer.

■ Question 3. Show me the handover, including at 7pm

- **A person is always reachable.** "Can I talk to someone" works as the first sentence, with no negotiation.
- **Context travels.** Whoever picks up already sees the number, the request, and what the agent could not do.
- **Frustration escalates.** Repetition, interruption, a second attempt. It gives up before the customer does.
- **Some subjects never enter it.** Complaints, legal, safety, vulnerable callers. A fixed list, not a judgement.
- **After hours means a booked callback** with a stated time you keep. An unanswered transfer at 7pm is worse than voicemail.

■ Question 4. Where does the conversation live?

Ask	Why it matters
Which country holds the audio and transcripts?	Often a hard requirement for health, legal, government, aged care and NDIS work.
Is my call content used to train a model?	You want a clear no in the contract, not a web page that can be edited.
How long is it kept, and can I set it?	Retention you cannot control is retention somebody else decided.
Who inside the provider can listen?	There is usually a legitimate support answer. There should also be a log.
How many companies are in the chain?	Each hop is another place data rests, and another party who can change terms. It is also the pause you can hear.

■ Start with these

- Hours, address, parking, what you cover.
- Booking, rescheduling, cancelling.
- Order, job and delivery status.
- After hours, with a real outcome.
- Daytime overflow.

■ Leave these alone

- Complaints of any kind.
- Distressed or vulnerable callers.
- Anything with a safety dimension.
- A cancellation you would fight to keep.
- The first call from a big new customer.

The honest reason to do it at all

It is rarely that a machine answers better than your best person. It is that at 8:40 on a Monday your best person is on the first call and cannot take the fourth, and the people who ring in that window mostly do not leave a message and do not ring back. They ring the next business on the list, and you never find out they existed. That invisible loss is what this addresses, which is why the number to watch is **total answered enquiries**, not call quality.

The number that lies

Containment rate is on every dashboard and flatters every failure: a caller who hung up in frustration is contained, and so is one told to ring back tomorrow. Watch resolution rate, repeat contact within 48 hours, escalation quality, abandonment inside the agent, and total answered enquiries against the same month last year. Then read twenty randomly chosen transcripts a week for the first two months. Random ones, not the ones the system flagged.